JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS **198,** 64–83 (1996)
ARTICLE NO. 0068

Iterative Parametric Minimax Method for a Class of Composite Optimization Problems

Duan Li*

Department of Systems Engineering and Engineering Management, Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

Jian-Bo Yang

Operational Research Group, School of Manufacturing and Mechanical Engineering, University of Birmingham, Birmingham B15 2TT, United Kingdom

Submitted by Koichi Mizukami

Received June 9, 1994

This paper considers a class of composite optimization problems that are often difficult to solve directly due to large dimension, nonlinearity, nonseparability, and/or nonconvexity of the problem. An iterative parametric minimax method is proposed in which the original optimization problem is embedded into a weighted minimax formulation. The resulting auxiliary parametric optimization problems at the lower level often have simple structures that are readily tackled by efficient solution strategies, such as the decomposition scheme in dynamic programming and in the primal-dual method. The analytical expression of the partial derivatives of systems performance indices with respect to the weighting vector in the parametric minimax formulation is derived. The gradient method can be thus adopted at the upper level to adjust the value of the weighting vector. The solution of the weighted minimax formulation converges to the optimal solution of the original problem in a multilevel iteration process. An application of the proposed iterative parametric minimax method is demonstrated in constrained reliability optimization problems. © 1996 Academic Press, Inc.

1. INTRODUCTION

We consider in this paper the following class of optimization problems where the overall objective function, J, is a composite function of multiple

* E-mail address: dli@se.cuhk.hk.

0022-247X/96 $18.00

systems performance indices, J_i $(i = 1, 2, \ldots, k)$,

$$\min J = \phi[J_1(x), J_2(x), \ldots, J_k(x)] \tag{1a}$$

subject to

$$g_j(x) \leq 0 \qquad j = 1, 2, \ldots, m \tag{1b}$$

with J being a nondecreasing function of J_i $(i = 1, 2, \ldots, k)$

$$\frac{\partial J}{\partial J_i} \geq 0 \tag{2}$$

where $x \in R^n$ is the decision vector and functions ϕ, J_i $(i = 1, 2, \ldots, k)$, and g_j $(j = 1, 2, \ldots, m)$ are all assumed to be second-order differentiable functions. Each function J_i $(i = 1, 2, \ldots, k)$ is assumed to possess a finite minimum value under constraint (1b). Without loss of generality, the minimum value of each J_i is assumed to be strictly positive. Furthermore, it is assumed that at the optimal point of problem (1), the corresponding value of each J_i $(i = 1, 2, \ldots, k)$ is finite.

This problem formulation is of wide application in optimization and control. One exemplary subject is in multiobjective optimization, where J_i $(i = 1, 2, \ldots, k)$ can be viewed as multiple systems attributes while J serves as the decisionmaker's disutility function. Another exemplary subject is the constrained reliability optimization problem, where J_i $(i = 1, 2, \ldots, k)$ represent unreliabilities of components (or subsystems) while J is the unreliability function of the overall network. The interpretation of (2) is that an improvement in each individual systems performance index leads to an improvement of the overall objective function. This assumption is well judged by many real-world optimization problems. A disutility function is a nondecreasing function of each system's attribute and the network unreliability is nondecreasing with respect to subsystems' unreliabilities when a network is of a coherent structure.

In many cases, the constrained optimization problem in (1) may be difficult to solve directly. Contributing factors to this difficulty could be a large dimension of such a problem, nonlinearity or nonseparability of the overall objective function with respect to systems performance indices, and/or nonconvexity of the problem. The motivation of this research is to establish a solution framework in which a composite optimization problem specified in (1) could be embedded into a family of parameterized problems which are much easier to be solved using existing efficient solution algorithms. More specifically, an efficient solution scheme is investigated in this paper, where the optimal solution of problem (1) is sought iteratively through a parametric minimax problem formulation.

Considerable research efforts have been reported in the literature in solving optimization problems using parametric solution procedures. The classical prime-dual method [1] leads in many cases to a decomposition of the original optimization problem. The reach of the primal-dual method was extended in [2] to nonconvex situations through certain convexification procedures. A weighted dynamic programming solution procedure was proposed in [3] for a discrete optimization problem with a quasiconcave or quasiconvex nonlinear utility function of the form $U(\sum f_t^1, \sum f_t^2)$. The C-programming method was developed in [4] for a class of nonseparable optimization problems whose performance indices are of the form of $\sum f_t(x_t) + \phi[\sum g_t(x_t)]$, where ϕ is either quasiconcave or quasiconvex. Multilevel parametric solution algorithms [5] were investigated for nonseparable dynamic programming problems [6–8], large-scale nonseparable optimization problems [9, 10], and general multiple linear-quadratic control problems [9, 11].

The organization of this paper is as follows. In Section 2, an iterative parametric minimax solution scheme is developed. It has been proved that certain optimal solutions of problem (1) can always be generated by a weighted minimax formulation. The analytical expression of the partial derivative of systems performance indices with respect to the weighting vector in the parametric minimax formulation is derived. A two-level solution algorithm is then proposed. At the lower level, the weighted minimax formulation is solved for a given weighting vector using a suitable solution algorithm. At the upper level, the weighting vector is adjusted using a gradient-type algorithm. This iteration process continues until the optimum stopping condition is met. Section 3 demonstrates an application of the proposed iterative parametric minimax method in constrained reliability optimization problems. The paper concludes in Section 4 with suggestions for future extensions.

2. ITERATIVE PARAMETRIC MINIMAX SOLUTION SCHEME

The motivation to develop a multilevel solution scheme is to embed a difficult constrained optimization problem in (1) into a family of parameterized optimization problems that are much easier to solve by existing efficient solution algorithms. The solution scheme should be devised such that the solution of the parametric problem converges to the optimum point of the original problem through successively adjusting the parameter vector in the iteration process.

Consider the following weighted minimax formulation for problem (1):

$$\min \max\{w_1 J_1(x), w_2 J_2(x), \ldots, w_k J_k(x)\} \tag{3a}$$

subject to

$$g_j(x) \leq 0 \qquad j = 1, 2, \ldots, m, \tag{3b}$$

where weighting coefficient w_1 is always set to one and weighting coefficients w_i $(i = 2, 3, \ldots, k)$ are nonnegative. In (3) the maximization is performed among k weighted systems performance indices while minimization is carried out over the feasible region of x. Denote weighting vector $[1, w_2, w_3, \ldots, w_k]$ by w and set $\{[1, w_2, w_3, \ldots, w_k] \mid w_i \geq 0,\ i = 2, 3, \ldots, k\}$ by W. Note here that the order of the systems performance indices is arbitrary.

Define X^* to be the solution set of problem (1), i.e.,

$$X^* = \{x \mid x \text{ is a minimizer of problem (1)}\}, \tag{4}$$

and X_w^* to be the union of the solutions sets of the weighted minimax problem in (3), i.e.,

$$X_w^* = \{x \mid x \text{ is a minimizer of problem (3) for a weighting vector } w \in W\}. \tag{5}$$

THEOREM 1. *The intersection of X^* and X_w^* is always nonempty,*

$$X^* \cap X_w^* \neq \varnothing \tag{6}$$

where $\varnothing$ is a null set.

Proof. If x^* is an optimal solution of problem (1), i.e., $x^* \in X^*$, set w_i^* equal to $J_1(x^*)/J_i(x^*)$ $(i = 2, 3, \ldots, k)$. If x^* is optimal in problem (3) with the constructed weighting vector, $w^* = [1, w_2^*, w_3^*, \ldots, w_k^*]$, then $x^* \in X_w^*$. Otherwise there exists a feasible solution $\hat{x}$ such that $\hat{x}$ is a minimizer of the weighted minimax problem with weighting vector equal to w^* and

$$\begin{aligned}
&\max\{J_1(\hat{x}), w_2^* J_2(\hat{x}), \ldots, w_k^* J_k(\hat{x})\} \\
&\qquad < \max\{J_1(x^*), w_2^* J_2(x^*), \ldots, w_k^* J_k(x^*)\} = J_1(x^*),
\end{aligned} \tag{7a}$$

i.e.,

$$J_i(\hat{x}) < J_i(x^*) \qquad \forall\ i = 1, 2, \ldots, k \tag{7b}$$

Since the overall objective function J is a nondecreasing function of J_i ($i = 1, 2, \ldots, k$), Eq. (7b) would lead to $J(\hat{x}) \le J(x^*)$. If the strict inequality holds, that will be a contradiction to the assumption of $x^* \in X^*$. If the equality holds, then we have $\hat{x} \in X_w^*$ and $\hat{x} \in X^*$. Q.E.D.

The implication of Theorem 1 is clear. If problem (1) has a unique solution, then this optimal solution can be generated by the weighted minimax formulation in (3). If problem (1) has multiple solutions, then at least a nonempty subset of X^* can be identified using the weighted minimax formulation in (3). Theorem 1 thus enables us to confine the search process for the optimal solution of problem (1) in the solution set of the weighted minimax formulation.

The minimax formulation given in (3) can be rewritten in the following equivalent form by introducing an auxiliary variable y:

$$\min \varphi(y) \tag{8a}$$

subject to

$$w_i J_i(x) \le y \qquad i = 1, 2, \ldots, k \tag{8b}$$

$$g_j(x) \le 0 \qquad j = 1, 2, \ldots, m \tag{8c}$$

where $\varphi(y)$ is any strictly increasing and second-order differentiable function of y. The equivalence of formulations (3) and (8) is evident. Each weighted systems performance index, $w_i J_i$, is bounded by y from above. Thus, minimizing a strictly increasing function of y will minimize the maximum value from among $w_1 J_1, w_2 J_2, \ldots, w_k J_k$. The simplest form of φ is of course y itself. The reason for adopting a general form of $\varphi(y)$, instead, is that the linearity of y in a formulation often leads to nonsatisfaction of the second-order sufficient conditions [12] at the optimal point of problem (8). The need to satisfy the second-order sufficient conditions will become clear in the following discussion, especially in Lemma 1. Note that y is always strictly positive since y is bounded by J_1 from below and the minimum point of J_1 is strictly positive.

Write the corresponding Lagrangian of problem (8) as

$$L_1 = \varphi(y) + \sum_{i=1}^{k} \lambda_i [w_i J_i(x) - y] + \sum_{j=1}^{m} \mu_j g_j(x) \tag{9}$$

where λ_i ($i = 1, 2, \ldots, k$) and μ_j ($j = 1, 2, \ldots, m$) are nonnegative Kuhn–Tucker multipliers. The set of the first-order necessary conditions for the optimality is

$$\frac{\partial L_1}{\partial y} = \frac{d\varphi(y)}{dy} - \sum_{i=1}^{k} \lambda_i = 0 \tag{10a}$$

$$\frac{\partial L_1}{\partial x} = \sum_{i=1}^{k} \lambda_i w_i \frac{\partial J_i(x)}{\partial x} + \sum_{j=1}^{m} \mu_j \frac{\partial g_j(x)}{\partial x} = 0 \tag{10b}$$

$$\lambda_i [w_i J_i(x) - y] = 0 \qquad i = 1, 2, \dots, k \tag{10c}$$

$$w_i J_i(x) - y \le 0 \qquad i = 1, 2, \dots, k \tag{10d}$$

$$\mu_j g_j(x) = 0 \qquad j = 1, 2, \dots, m \tag{10e}$$

$$g_j(x) \le 0 \qquad j = 1, 2, \dots, m. \tag{10f}$$

One important observation is that if the Kuhn–Tucker multipliers, λ_i $(i = 1, 2, \dots, k)$, are all strictly positive the optimization process of a minimax formulation acts as an equalizer to make all $w_i J_i$ equal to y.

Problem (8) can be rewritten as

$$\min \varphi(y) \tag{11a}$$

subject to

$$-\frac{y}{J_i(x)} \le -w_i \qquad i = 1, 2, \dots, k \tag{11b}$$

$$g_j(x) \le 0 \qquad j = 1, 2, \dots, m. \tag{11c}$$

LEMMA 1. *Assume that the optimal solution* (x^*, y^*) *of problem* (8) (*and problem* (11)) *is a regular point* [12]. *In addition,* (x^*, y^*), *together with the associated Kuhn–Tucker multipliers* σ_i $(i = 1, 2, \dots, k)$ *and* γ_j $(j = 1, 2, \dots, m)$, *satisfies the second-order sufficient conditions of problem* (11), *where* σ_i $(i = 1, 2, \dots, k)$ *and* γ_j $(j = 1, 2, \dots, m)$ *are nonnegative Kuhn–Tucker multipliers associated with Eqs*. (11b) *and* (11c), *respectively*. *Assume further that all active inequalities in problem* (8) (*and problem* (11)) *are nondegenerate* [12]. *Then we have*

$$\frac{\partial \varphi(y^*)}{\partial w_i} = \lambda_i J_i(x^*). \tag{12}$$

Proof. Write the corresponding Lagrangian of problem (11) as

$$L_2 = \varphi(y) + \sum_{i=1}^{k} \sigma_i [-y/J_i(x) + w_i] + \sum_{j=1}^{m} \gamma_j g_j(x). \tag{13}$$

The set of the first-order necessary conditions for the optimality is

$$\frac{\partial L_2}{\partial y} = \frac{d\varphi(y)}{dy} - \sum_{i=1}^{k} \sigma_i / J_i(x) = 0 \tag{14a}$$

$$\frac{\partial L_2}{\partial x} = \sum_{i=1}^{k} \sigma_i \frac{y}{J_i^2(x)} \frac{\partial J_i(x)}{\partial x} + \sum_{j=1}^{m} \gamma_j \frac{\partial g_j(x)}{\partial x} = 0 \tag{14b}$$

$$\sigma_i[-y/J_i(x) + w_i] = 0 \quad i = 1, 2, \ldots, k \tag{14c}$$

$$-y/J_i(x) + w_i \leq 0 \qquad i = 1, 2, \ldots, k \tag{14d}$$

$$\gamma_j g_j(x) = 0 \qquad j = 1, 2, \ldots, m \tag{14e}$$

$$g_j(x) \leq 0 \qquad j = 1, 2, \ldots, m \tag{14f}$$

Comparing the two sets of necessary conditions in Eqs. (10) and (14), it is evident that $\{x^*, y^*, \lambda_i\ (i = 1, 2, \ldots, k),\ \mu_j\ (j = 1, 2, \ldots, m)\}$ solves Eq. (10) if and only if $\{x^*, y^*, \sigma_i = \lambda_i J_i(x^*)\ (i = 1, 2, \ldots, k),\ \gamma_j = \mu_j\ (j = 1, 2, \ldots, m)\}$ solves Eq. (14). Note here that (i) if the ith constraint in Eq. (8b) (and in Eq. (11b)) is not binding, both λ_i and σ_i are equal to zero, and (ii) if the ith constraint in Eq. (8b) (and in Eq. (11b)) is binding, y^* is equal to $w_i J_i(x^*)$ and both λ_i and σ_i are nonzero due to the nondegenerate assumption.

From the sensitivity theorem [12], we know that $\partial\varphi(y^*)/\partial(-w_i) = -\sigma_i$ in problem (11) if the conditions stated in the theorem are satisfied, i.e.,

$$\frac{\partial\varphi(y^*)}{\partial w_i} = \sigma_i \tag{15}$$

Thus, we further have

$$\frac{\partial\varphi(y^*)}{\partial w_i} = \lambda_i J_i(x^*) \tag{16}$$

Q.E.D.

Optimal solution of problem (8) depends on the assigned value of w. In abstract, the optimal solution of problem (8) can be parameterized by the weighting vector w,

$$x^* = x^*(w) \tag{17a}$$

$$y^* = y^*(w) \tag{17b}$$

$$\lambda_i^* = \lambda_i^*(w) \qquad i = 1, 2, \ldots, k \tag{17c}$$

$$\mu_j^* = \mu_j^*(w) \qquad j = 1, 2, \ldots, m \tag{17d}$$

$$J_i^* = J_i^*(w) \qquad i = 1, 2, \ldots, k. \tag{17e}$$

It is assumed in this paper that all functions in (17) are differential with respect to w. In the case where the search can be carried out in the neighborhood of w^* with which the optimal solution of (8) attains the optimum point of (1), this assumption can be relaxed to that all functions in (17) are differential with respect to w in the neighborhood of w^*. Substituting $J_i^*(w)$ into overall objective function J, J becomes a function of w,

$$J^* = J^*(w). \tag{18}$$

From Theorem 1, the optimal point of J can be achieved through minimizing $J^*(w)$ with respect to w. The key task in the following is to derive the derivative of J with respect to w at the current solution of the weighted minimax formulation. If problem (8) can be solved analytically, the function forms in Eq. (17e) will be ready to hand and the gradient of J with respect to each w_i is easily made available. In most real situations, however, problem (8) is solved numerically. There will be then no knowledge about the function form in Eq. (17e). It is thus necessary to derive the derivatives using only current local information of pointwise values of the decision variables and the Kuhn–Tucker multipliers. After the derivative of J with respect to w is obtained, the search for the optimal point of $J(w)$ can be implemented using the gradient method. The search for the optimal point of $J(w)$ then is an unconstrained problem except w_i's are bounded to be nonnegative.

THEOREM 2. *If, for a given weighting vector* w, *the optimal Kuhn–Tucker multiplier associated with the ith constraint in Eq.* (8b) *is strictly positive at the optimal solution*, $\{x^*, y^*, \lambda^*, \mu^*\}$, *of problem* (8), *then we have the following equality in the neighborhood of* $\{x^*, y^*, \lambda^*, \mu^*\}$,

$$\frac{\partial J_i^*}{\partial w_j} = \left[\frac{\lambda_j^* J_j^*}{d\varphi(y^*)/dy} - \delta(i-j) J_i^* \right] / w_i, \quad i = 1, 2, \dots, k;\ j = 2, 3, \dots, k \tag{19}$$

where $\delta(\cdot)$ *is the Kronecker function,* $\delta(0) = 1$, *and* $\delta(x) = 0$ *for all* $x \neq 0$.

Proof. If the optimal Kuhn–Tucker multiplier associated with the ith constraint in Eq. (8b) is strictly positive at $\{x^*, y^*, \lambda^*, \mu^*\}$, then we have the following from Eq. (10c) in the neighborhood of $\{x^*, y^*, \lambda^*, \mu^*\}$,

$$w_i J_i^*(w) - y^*(w) = 0. \tag{20}$$

Taking derivative with respect to w_j on both sides of Eq. (20) yields

$$\delta(i-j) J_i^* + w_i \frac{\partial J_i^*}{\partial w_j} - \frac{\partial y^*}{\partial w_j} = 0. \tag{21}$$

From Eq. (16), we have

$$\frac{\partial y^*}{\partial w_j} = \frac{\lambda_j^* J_j^*}{d\varphi(y^*)/dy}. \tag{22}$$

Substituting (22) into (21) and solving it for $\partial J_i^*/\partial w_j$, we prove the theorem. Q.E.D.

Note here that if the ith weighting coefficient w_i is zero, the ith constraint in Eq. (8b) will not be binding since y is strictly positive. Thus, when the optimal Kuhn–Tucker multiplier associated with the ith constraint in Eq. (8b) is strictly positive at the optimal solution, the corresponding w_i must be strictly positive.

If all the optimal Kuhn–Tucker multipliers associated with the constraints in Eq. (8b) are strictly positive at $\{x^*, y^*, \lambda^*, \mu^*\}$, then we have the following in the neighborhood of $\{x^*, y^*, \lambda^*, \mu^*\}$,

$$\frac{\partial J^*}{\partial w_i} = \sum_{j=1}^{k} \frac{\partial J^*}{\partial J_j^*} \frac{\partial J_j^*}{\partial w_i} \qquad i = 2, 3, \dots, k. \tag{23}$$

From Theorem 1, we know that if an optimal solution x^* of (1) belongs to the set X_w^* defined in (5), it can be always generated by the weighted minimax formulation in (3) with weighting vector $w^* = [1, J_1(x^*)/J_2(x^*), J_1(x^*)/J_3(x^*), \dots, J_1(x^*)/J_k(x^*)]$. Note here that each component of w^* is strictly positive due to the assumptions made in (1). Thus, the search for optimum weighting vector can be confined in $W_+ = \{[1, w_2, w_3, \dots, w_k] \mid w_i > 0,\ i = 2, 3, \dots, k\}$.

Assuming that all the assumptions in Lemma 1 are satisfied during the whole iteration process, the original composite optimization problem could be then solved in a two-level solution structure. For a given weighting vector w, weighted minimax problem (3) or (8) is solved at the lower level using appropriate solution schemes. Depending on the problem structure of the parametric minimax formulation in (3) or (8), linear programming, nonlinear programming, dynamic programming, or other efficient methods can be used as the solution algorithm at the lower level. If the solution scheme at the lower level does not furnish the corresponding Kuhn–Tucker multipliers, the Kuhn–Tucker multipliers can be found by solving the set of first-order Kuhn–Tucker conditions of problem (8) along with the available knowledge of the identified optimal solution of $\{x^*, y^*\}$. Note here that the set of first-order Kuhn–Tucker conditions of (8) is linear in Kuhn–Tucker multipliers when solution $\{x^*, y^*\}$ is known. At the upper level, the optimal stopping condition $\{\partial J^*/\partial w_i = 0,\ i = 2, 3, \dots, k\}$ is checked upon receiving the solutions from the lower level. If it is not satisfied, a gradient-type algorithm can be used to update the value of the weighting vector:

$$w_i^{t+1} = \max\left\{0, w_i^t - \alpha \frac{\partial J^*}{\partial w_i}\right\} \qquad i = 2, 3, \ldots, k, \tag{24}$$

where t is the iteration number and α is a step-size parameter which can be adjusted during the iteration to guarantee a decrement of overall objective function J. Problem (3) or (8) at the lower level is then solved again for this new value of w. The iteration process continues until all $(\partial J^*/\partial w_i)$'s vanish.

The overall algorithm of the iterative parametric minimax method is now summarized as follows.

ALGORITHM OF ITERATIVE PARAMETRIC MINIMAX METHOD.

Step 1. *Select initial weighting vector* w^0, *choose a very small number* ε *for the error tolerance in the stopping condition, and set iteration number* $t = 0$.

Step 2. *For the selected weighting vector* w^t, *solve problem* (3) *or* (8) *and obtain solution* $\{x^t, y^t, \lambda^t\}$.

Remark 1. Assume at iteration t, some inequalities in (8b) are not binding when a solution $\{x^t, y^t\}$ is generated at Step 2. The value of w^t needs to be reassessed and problem (8) needs to be solved again in order to calculate $\partial J_i^*/\partial w_j$. In the situations where $J_1(x^t) = y^t$, solution $\{x^t, y^t\}$ remains feasible and optimal in (8) with all constraints in (8b) binding when the inactive constraints in (8b) are modified to

$$\frac{y^t}{J_i(x^t)} J_i(x) \le y \qquad \text{if } J_i(x^t) < y^t, i \in \{2, 3, \ldots, k\}. \tag{25}$$

In the situations where $J_1(x^t) < y^t$, solution $\{x^t, y^t\}$ remains feasible and optimal in (8) with all constraints in (8b) binding when the inactive constraints in (8b) are modified to

$$\frac{y^t}{J_1(x^t)} J_1(x) \le y \tag{26a}$$

$$\frac{y^t}{J_i(x^t)} J_i(x) \le y \qquad \text{if } J_1(x^t) < y^t, \qquad i \in \{2, 3, \ldots, k\} \tag{26b}$$

Normalizing the weighting coefficients specified in (26), the following reassessment scheme can be developed for both situations where either

$J_1(x^t) = y^t$ or $J_1(x^t) < y^t$. Denote w^{t-M} as the modified value of w^t at iteration t. The value of w^{t-M} is set according to the following formula

$$w_i^{t-M} = \frac{J_1(x^t)}{J_i(x^t)} \qquad i = 2, 3, \ldots, k. \tag{27}$$

Reassessing the weighting vector according to (27) will keep x^t as the optimal solution at iteration t and at the same time make all inequalities in (8b) binding. Problem (8) can be then solved again using w^{t-M} to obtain the necessary information in order to calculate $\partial J_i^* / \partial w_j$ at iteration t.

Remark 2. If the solution procedure at Step 2 does not provide corresponding Kuhn–Tucker multipliers, solve λ^t and μ^t from the first-order Kuhn–Tucker conditions of (8) with x and y being set to x^t and y^t, respectively.

Step 3. Check if a norm of $\partial J^* / \partial w$ is less than or equal to the preselected small number ε,

$$\left\| \frac{\partial J^*}{\partial w} \right\| \leq \varepsilon. \tag{28}$$

If yes, the search terminates. Otherwise go to Step 4.

Step 4. Update w using Eq. (24). Set $t = t + 1$ and go back to Step 2.

Two example problems will be presented in the following to illustrate the step-by-step procedure of the proposed solution algorithm.

Example 1. Consider the following nonlinear programming problem,

$$\min J = \left\{150 \exp(x_1 + x_2 + x_3) + \Sigma_{i=1}^3 (x_i + c_i)^2\right\} \tag{29a}$$

subject to

$$\sum_{i=1}^{3} \left[\exp(a_i x_i) + b_i x_i^2\right] \leq 10 \tag{29b}$$

$$x_1, x_2, x_3 \leq 0 \tag{29c}$$

where $a_1 = 2$, $a_2 = 1$, $a_3 = 3$, $b_1 = 1$, $b_2 = 3$, $b_3 = 2$, $c_1 = 1$, $c_2 = 2$, and $c_3 = 3$. The above problem is highly nonlinear and nonseparable with respect to x_1, x_2, and x_3. Define

$$J_1 = 8 + x_1 + x_2 + x_3 \tag{30a}$$

$$J_2 = \sum_{i=1}^{3} (x_i + c_i)^2 \tag{30b}$$

Objective J can be expressed as a function of J_1 and J_2,

$$J = 150 \exp(J_1 - 8) + J_2 \tag{31}$$

with both $\partial J / \partial J_1$ and $\partial J / \partial J_2$ being strictly positive. The reason to include a constant term 8 in J_1 is to guarantee J_1 to be strictly positive. From (3), the following weighted minimax problem is formulated

$$\min \max \left\{ 8 + x_1 + x_2 + x_3, w_2 \sum_{i=1}^{3} (x_i + c_i)^2 \right\} \tag{32}$$

subject to Eqs. (29b) and (29c).

Choosing $\varphi(y)$ to be y^2, the following equivalent problem can be formed

$$\min y^2 \tag{33a}$$

subject to

$$8 + x_1 + x_2 + x_3 \leq y \tag{33b}$$

$$w_2 \sum_{i=1}^{3} (x_i + c_i)^2 \leq y \tag{33c}$$

and Eqs. (29b) and (29c).

Problem (33) is convex and separable with respect to y, x_1, x_2, and x_3 and can be solved efficiently using the primal-dual method. The dual function of (33) is

$$\begin{aligned} H(\lambda_1, \lambda_2, \mu) = \min \Bigg\{ & y^2 + \lambda_1 [8 + x_1 + x_2 + x_3 - y] \\ & + \lambda_2 \left[w_2 \sum_{i=1}^{3} (x_i + c_i)^2 - y \right] \\ & + \mu \left[\sum_{i=1}^{3} \left[\exp(a_i x_i) + b_i x_i^2 \right] - 10 \right] \Bigg\}. \end{aligned} \tag{34}$$

For given values of λ_1, λ_2, and μ, the above dual function can be solved through decomposition at the lower level,

Subproblem 0: Solving

$$\min y^2 - \lambda_1 y - \lambda_2 y \tag{35a}$$

subject to

$$y \geq 0 \tag{35b}$$

yields the following optimal solution

$$y^* = (\lambda_1 + \lambda_2)/2. \tag{36}$$

Subproblem i ($i = 1, 2, 3$): Solving

$$\min\left\{\lambda_1 x_i + \lambda_2 w_2 (x_i + c_i)^2 + \mu\left[\exp(a_i x_i) + b_i x_i^2\right]\right\} \tag{37a}$$

subject to

$$x_i \le 0 \tag{37b}$$

yields optimal solution x_i^* that satisfies

$$\mu a_i \exp(a_i x_i^*) = -(\lambda_1 + 2\lambda_2 w_2 c_i) - 2(\lambda_2 w_2 + \mu b_i) x_i^*. \tag{38}$$

Denote the iteration number in the primal-dual approach by s. The values of λ_1, λ_2, and μ are adjusted at the second level by maximizing the dual function

$$\lambda_1^{s+1} = \max\{0, \lambda_1^s + \alpha_2[8 + x_1 + x_2 + x_3 - y]\} \tag{39a}$$

$$\lambda_2^{s+1} = \max\left\{0, \lambda_2^s + \alpha_2\left[w_2 \sum_{i=1}^{3} (x_i + c_i)^2 - y\right]\right\} \tag{39b}$$

$$\mu^{s+1} = \max\left\{0, \mu^s + \alpha_2\left[\sum_{i=1}^{3}\left[\exp(a_i x_i) + b_i x_i^2\right] - 10\right]\right\} \tag{39c}$$

where α_2 is a step-size parameter which can be adjusted on-line to guarantee an increment of the dual function. The primal-dual solution process in solving (33) continues until the optimal conditions $\lambda_1(\partial H/\partial\lambda_1) = 0$, $\lambda_2(\partial H/\partial\lambda_2) = 0$, and $\mu(\partial H/\partial\mu) = 0$ are met.

Each time after the solution of (33) is obtained for a given weighting coefficient w_2, a new value of w_2 is calculated at the third level using (24),

$$w_2^{t+1} = w_2^t - \alpha_3 \frac{\partial J}{\partial w_2}, \tag{40}$$

where α_3 is a step-size parameter which can be adjusted on-line to guarantee a decrement of function J. Derivative $\partial J/\partial w_2$ is evaluated using (23),

$$\frac{\partial J}{\partial w_2} = 150 \exp(J_1 - 8)(\partial J_1/\partial w_2) + (\partial J_2/\partial w_2) \tag{41}$$

where $(\partial J_1/\partial w_2)$ and $(\partial J_2/\partial w_2)$ are obtained using (19),

$$\frac{\partial J_1}{\partial w_2} = \lambda_2 \sum_{i=1}^{3} (x_i + c_i)^2/(2y) \tag{42a}$$

$$\frac{\partial J_2}{\partial w_2} = \left[\lambda_2 \sum_{i=1}^{3} (x_i + c_i)^2/(2y) - \sum_{i=1}^{3} (x_i + c_i)^2\right]/w_2. \tag{42b}$$

The initial value of w_2 is set to 1 and the step size parameters α_2 and α_3 are selected to be equal to 0.1 and 0.1, respectively. The stopping criterion is that the absolute value of $\partial J^*/\partial w_2$ is less than 0.0001. The iteration process converges very fast and it ends at the ninth iteration with optimal solution $w_2 = 1.279266$, $x_1 = -1.339983$, $x_2 = -0.9676399$, $x_3 = -1.571332$, and $J = 6.323308$.

The advantage of adopting the proposed iterative parametric minimax method in Example 1 is evident. The original nonseparable constrained optimization problem is embedded into a family of auxiliary weighted minimax formulation that is of a convex and separable structure. The primal-dual method can be then applied. The resulting decomposition in solving weighted minimax formulation greatly reduces the problem complexity, while the search for optimal w_2 value at the upper level is only one dimensional with a simple lower bound.

EXAMPLE 2. Consider the following nonlinear programming problem,

$$\min J = \left\{\exp(3x_1) + \exp(4x_2) + x_1(x_2)^2\right\} \tag{43a}$$

subject to

$$(x_1)^2 + 2(x_2)^2 \geq 1 \tag{43b}$$

$$x_1 \geq 0, \qquad x_2 \geq 0 \tag{43c}$$

The above problem is nonconvex and highly nonlinear. It is obvious that J is an increasing function of both x_1 and x_2. Let J_1 be defined as x_1 and J_2 as x_2. The following weighted minimax problem is formulated

$$\min \max\{x_1, w_2 x_2\} \tag{44}$$

subject to Eqs. (43b) and (43c).

Choosing $\varphi(y)$ to be y^2, the following problem is equivalent to problem (44)

$$\min y^2 \tag{45a}$$

subject to

$$x_1 \leq y \tag{45b}$$

$$w_2 x_2 \leq y, \tag{45c}$$

and Eqs. (43b) and (43c).

The analytical solutions of problem (43) can be found without much difficulty,

$$y = x_1 = \frac{w_2}{\sqrt{2 + w_2^2}} \tag{46a}$$

$$x_2 = \frac{1}{\sqrt{2 + w_2^2}} \tag{46b}$$

$$\lambda_1 = \frac{2w_2^3}{\left(2 + w_2^2\right)^{1.5}} \tag{46c}$$

$$\lambda_2 = \frac{4w_2}{\left(2 + w_2^2\right)^{1.5}}. \tag{46d}$$

The derivatives of x_1 and x_2 with respect to w_2 can be calculated using (19). The results can be also verified by taking derivatives directly in (46a) and (46b) since the analytical solution is available in this example.

$$\frac{\partial x_1}{\partial w_2} = \frac{\lambda_2 x_2}{2y} = \frac{2}{\left(2 + w_2^2\right)^{3/2}} \tag{47a}$$

$$\frac{\partial x_2}{\partial w_2} = \frac{(\lambda_2 x_2/2y) - x_2}{w_2} = \frac{-w_2}{\left(2 + w_2^2\right)^{3/2}} \tag{47b}$$

The derivative of J with respect to w_2 can be now expressed as

$$\frac{\partial J}{\partial w_2} = \left(3\exp(3x_1) + x_2^2\right)\frac{\partial x_1}{\partial w_2} + \left(4\exp(4x_2) + 2x_1x_2\right)\frac{\partial x_2}{\partial w_2}. \tag{48}$$

The initial value of w_2 is set to 2 and the step size parameter α in Eq. (24) is selected to be equal to 0.1. The stopping criterion is that the absolute value of $\partial J^*/\partial w_2$ is less than 0.00001. The iteration process converges very fast and it ends at the 20th iteration with optimal solution $w_2 = 1.198401$, $x_1 = 0.646495$, $x_2 = 0.539465$, and $J = 15.7959$.

The advantage of adopting the proposed iterative parametric minimax method in Example 2 is clear. The original nonlinear constrained optimization problem is solved by a two level structure in which the first level can be solved analytically while the second level only involves a one-dimensional search with a simple one-side bound.

Some prominent features of the proposed iterative parametric minimax method need to be emphasized. The iterative parametric minimax method applies to very general situations. Many specific assumptions, such as the convexity, are not required. There exists great flexibility in choosing the forms of J_i $(i = 1, 2, \ldots, k)$. A thoughtful selection could significantly facilitate the solution process, as witnessed in the above two examples.

3. APPLICATION IN CONSTRAINED RELIABILITY OPTIMIZATION

The proposed iterative parametric minimax method is applied in this section to a class of constrained reliability optimization problems where the network is of a coherent structure [13] and consists of k components. The unreliability of a network, Q, can be expressed as a function of unreliabilities of k components, $\phi(q_1, q_2, \ldots, q_k)$. The objective is to minimize the network unreliability under a resource constraint:

$$\min Q = \phi(q_1, q_2, \ldots, q_k) \tag{49a}$$

subject to

$$\sum_{i=1}^{k} C_i(q_i) \leq C \tag{49b}$$

$$0 \leq L_i \leq q_i \leq U_i \leq 1, \qquad i = 1, 2, \ldots, k, \tag{49c}$$

where $C_i(q_i)$ represents the amount of the resource consumed by the ith component with unreliability q_i, C is the total amount of the resource, L_i is the minimum value of unreliability which the ith component can achieve, and U_i is the maximum value of unreliability for the ith component which is not allowed to exceed. It is well known that for a network of coherent structure, the network unreliability is a nondecreasing function of the unreliability of each component, i.e.,

$$\frac{\partial Q}{\partial q_i} \geq 0 \qquad i = 1, 2, \ldots, k. \tag{50}$$

It is clear that the reliability optimization problem in Eq. (49) is a special case of the composite optimization problem posed in (1). In this special case, each J_i in (1) simply reduces to the ith decision variable, q_i, and g_j in (1) has an additive form with respect to $C_i(q_i)$ $(i = 1, 2, \ldots, k)$. The usefulness of the proposed iterative parametric minimax method in constrained reliability optimization will be demonstrated through the following example problem.

EXAMPLE 3. Consider a variant of a reliability optimization problem in Hwang *et al.* [14] and Li and Haimes [15]. The unreliability, Q, of the network shown in Fig. 1 is to be minimized under a cost constraint,

$$\min Q = \left\{ (q_1)^2 (1 - q_3)(q_4)^2 + q_3 \left[1 - (1 - q_2)(1 - q_1 q_4) \right]^2 \right\} \quad (51a)$$

subject to

$$\sum_{i=1}^{4} \frac{G_i}{q_i} \leq 1, \quad (51b)$$

where $G_1 = 0.03$, $G_2 = 0.03$, $G_3 = 0.06$, and $G_4 = 0.04$. The term G_i/q_i represents the cost associated with q_i. The smaller the value of q_i, the larger the corresponding cost. This problem is highly nonlinear and non-separable in the sense of dynamic programming.

Introducing a state variable s_i as the sum of the costs for components 1 to $i - 1$ with s_1 equal to zero, the following auxiliary weighted minimax problem is formulated as in (3):

$$\min \max \{ w_1 q_1, w_2 q_2, w_3 q_3, w_4 q_4 \} \quad (52a)$$

subject to

$$s_{i+1} = s_i + G_i/q_i \qquad i = 1, 2, 3, 4 \quad (52b)$$

$$s_1 = 0, \qquad s_5 \leq 1$$

where w_1 is set to one. The above minimax problem can be solved using dynamic programming with the analytical solution

$$q_i = \frac{P_i}{w_i(1 - s_i)}, \qquad i = 1, 2, 3, 4, \quad (53)$$

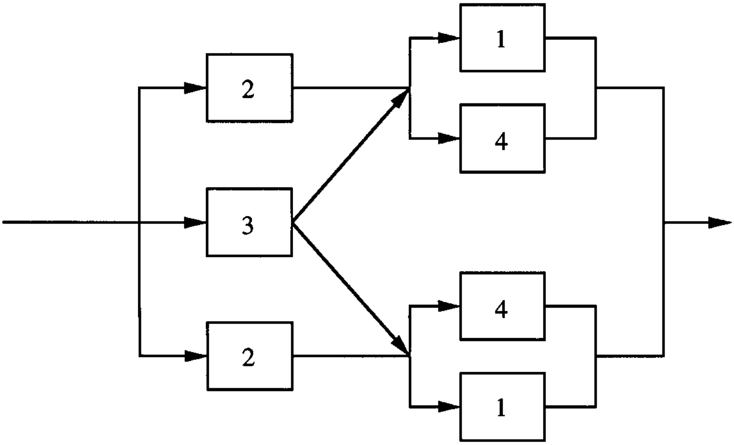

FIG. 1. Schematic diagram of a reliability optimization problem.

where P_i is calculated backward from $i = 3$ to $i = 1$,

$$P_i = P_{i+1} + w_i G_i, \qquad i = 3, 2, 1, \tag{54}$$

with

$$P_4 = w_4 G_4. \tag{55}$$

Selecting $\varphi(y)$ to be y^2, the corresponding Kuhn−Tucker multipliers associated with Eq. (8b) can be obtained by solving a set of first-order Kuhn−Tucker conditions that are linear equations with respect to y, λ_i $(i = 1, 2, 3, 4)$, and μ after the values of q_i $(i = 1, 2, 3, 4)$ are generated using dynamic programming,

$$\lambda_i = 2G_i q_1 \Big/ \left[w_i (q_i)^2 \left(\sum_{j=1}^{4} \frac{G_j}{w_j (q_j)^2} \right) \right] \qquad i = 1, 2, 3, 4 \tag{56}$$

Using Eq. (19), the partial derivatives of q_i with respect to w_j can be derived using the formulas

$$\frac{\partial q_i}{\partial w_j} = \frac{\lambda_j q_j}{2 w_i q_1} \qquad i = 1, 2, 3, 4;\ j = 2, 3, 4;\ i \neq j \tag{57a}$$

$$\frac{\partial q_i}{\partial w_i} = \left(\frac{\lambda_i q_i}{2 q_1} - q_i \right) \Big/ w_i \qquad i = 2, 3, 4. \tag{57b}$$

The partial derivatives of the network unreliability with respect to components' unreliabilities in this example are given as

$$\frac{\partial Q}{\partial q_1} = 2q_1(1 - q_3)(q_4)^2 + 2(1 - q_2)q_3 q_4 [1 - (1 - q_2)(1 - q_1 q_4)] \tag{58a}$$

$$\frac{\partial Q}{\partial q_2} = 2q_3(1 - q_1 q_4)[1 - (1 - q_2)(1 - q_1 q_4)] \tag{58b}$$

$$\frac{\partial Q}{\partial q_3} = [1 - (1 - q_2)(1 - q_1 q_4)]^2 - (q_1)^2 (q_4)^2 \tag{58c}$$

$$\frac{\partial Q}{\partial q_4} = 2(q_1)^2(1 - q_3)q_4 + 2q_1(1 - q_2)q_3[1 - (1 - q_2)(1 - q_1 q_4)]. \tag{58d}$$

The upper level adjusts the value of the weighting vector using (24),

$$w_i^{t+1} = w_i^t - \alpha \frac{\partial Q}{\partial w_i}$$

$$= w_i^t - \alpha \sum_{j=1}^{4} \frac{\partial Q}{\partial q_j} \frac{\partial q_j}{\partial w_i}, \qquad i = 2, 3, 4, \tag{59}$$

where α is set to be equal to 100 in this example problem.

The initial values of weighting coefficients are all set to one. The stopping criterion is $\sum_{i=2}^{4}(\partial Q/\partial w_i)^2 \leq 10^{-10}$. The search process terminates at the 27th iteration when $w_2 = 1.73830$, $w_3 = 0.54828$, and $w_4 = 0.74398$. The resulting optimal solution is $q_1 = 0.14480$, $q_2 = 0.08330$, $q_3 = 0.26411$, and $q_4 = 0.19464$. The corresponding network unreliability is 0.00373043.

4. CONCLUSIONS

An iterative parametric minimax method is proposed in this paper for a class of composite optimization problems. The systems' performance indices in a composite objective function can either be defined in the original problem description or be artificially introduced for solution convenience. The proposed approach tackles difficult and complex optimization problems through embedding, separation, decomposition, and coordination. The resulting auxiliary parametric optimization problems at the lower level are of a minimax type and often have a simple structure that is mathematically tractable by existing efficient solution algorithms. The analytical expression of the partial derivatives of systems performance indices with respect to the weighting vector is derived. This guarantees the convergence of a gradient-type search algorithm in adjusting the weighting vector at the upper level.

Promising results in this paper call for further exploration fo a nontraditional research direction using solution scheme of parameterization. An extension to discrete composite optimization will be of a great significance in application areas such as the constrained redundancy optimization problems in networks.

REFERENCES

1. L. S. Lasdon, "Optimization Theory for Large Systems," Macmillan, London, 1970.
2. D. P. Bertsekas, "Constrained Optimization and Lagrange Multiplier Methods," Academic Press, Boston, 1982.

3. M. Henig, The shortest path problem with two objective functions, *Europ. J. Oper. Res.* **25** (1985), 281–291.
4. M. Sniedovich, ''Dynamic Programming,'' Dekker, New York, 1992.
5. D. Li and Y. Y. Haimes, Using multiobjective optimization as a separation strategy for nonseparable problems, *in* ''Multiple Criteria Decision Making'' (G.-H. Tzeng, H.-F. Wang, U.-P. Wen, and P. L. Yu, Eds.), Springer-Verlag, New York, 1994.
6. D. Li, Multiple objectives and nonseparability in stochastic dynamic programming, *Internat. J. Systems Sci.* **21** (1990), 933–950.
7. D. Li and Y. Y. Haimes, New approach for nonseparable dynamic programming problems, *J. Optim. Theory Appl.* **64** (1990), 311–330.
8. D. Li and Y. Y. Haimes, Extension of dynamic programming to nonseparable problems, *Computer Math. Appl.* **21** (1991), 51–56.
9. D. Li, Hierarchical control for large-scale systems with general multiple linear-quadratic structure, *Automatica* **29** (1993), 1451–1461.
10. D. Li and Y. Y. Haimes, Multilevel methodology for a class of nonseparable optimization problems, *Internat.J. Systems Sci.* **21** (1990), 2351–2360.
11. D. Li, On general multiple linear-quadratic control problems, *IEEE Trans. Automatic Control* **38** (1993), 1722–1727.
12. D. G. Luenberger, ''Linear and Nonlinear Programming,'' 2nd ed., Addison–Wesley, Reading, MA, 1984.
13. A. Kaufmann, D. Grouchko, and R. Cruon, ''Mathematical Models for the Study of the Reliability of Systems,'' Academic Press, New York, 1977.
14. C. L. Hwang, F. A. Tillman, and W. Kuo, Reliability optimization by generalized lagrangian-function and reduced-gradient methods, *IEEE Trans. Reliability* **28** (1979), 316–319.
15. D. Li and Y. Y. Haimes, A decomposition method for optimization of large system reliability, *IEEE Trans. Reliability* **41** (1992), 183–189.